AD-A216 717

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

DTIC FILE COPY

(2)

DTIC ELECTE JAN 12 1990 S D D

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	Unclassified
1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
3. DISTRIBUTION/AVAILABILITY OF REPORT	Approved for public release; distribution unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	
5. MONITORING ORGANIZATION REPORT NUMBER(S)	AFOSR-TR- 89-1673
6a. NAME OF PERFORMING ORGANIZATION	Center for Research in Math and Science Education
6b. OFFICE SYMBOL (If applicable)	
7a. NAME OF MONITORING ORGANIZATION	Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code)	San Diego State University San Diego, CA 92182
7b. ADDRESS (City, State and ZIP Code)	Building 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION	AFOSR
8b. OFFICE SYMBOL (If applicable)	NL
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	AFOSR-89-0107
8c. ADDRESS (City, State and ZIP Code)	Building 410 Bolling AFB, DC 20332-6448

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2313	A4	

11. TITLE (Include Security Classification)	Schema-based Theories of Problem Solving
12. PERSONAL AUTHOR(S)	Stephen K. Reed
13a. TYPE OF REPORT	Annual Report
13b. TIME COVERED	FROM 11/1/88 TO 10/31/89
14. DATE OF REPORT (Yr., Mo., Day)	1989 November 1
15. PAGE COUNT	26
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES

FIELD	GROUP	SUB. GR.
05	09	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Schema, examples, procedures, analogs, categorization, word problems

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The objective of this research is to construct a schema-based model of problem solving to represent the construction of equations for solving algebra word problems. The research summarized in this report is concerned with the selection, use, and description of instructional examples Experiment 1 shows that mathematical experience was beneficial for improving the selection of good analogies when the analogies are isomorphic to the test problems, but was not beneficial when the analogies are more inclusive than the test problems. In Experiment 2 students were able to effeetively combine information from two analogous problems but did significantly worse when combining information from one example and a set of procedures. The last three experiments required that students categorize motion problems according to whether the two distances in a problem should be equated, added, or subtracted. (OVER)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☒ DTIC USERS ☐
21. ABSTRACT SECURITY CLASSIFICATION	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	Dr. Alfred R. Fregly
22b. TELEPHONE NUMBER (Include Area Code)	(202)-767-5021
22c. OFFICE SYMBOL	NL

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

90 01 11 107

Categorization significantly improved as the number of training examples representing a category increased from one to four (Experiment 3). Categorization was also significantly better when students received both specific and general descriptions of the examples than when they received only a single description (Experiment 4). However, as shown in Experiment 5, students were unable to form their own general descriptions by comparing similar examples.

Table of Contents

AFOSR-TR- 89-1673

SUMMARY

The objective of this research is to construct a schema-based model of problem solving to represent the construction of equations for solving algebra word problems. The research summarized in this report is concerned with the selection, use, and description of instructional examples. Experiment 1 shows that mathematical experience was beneficial for improving the selection of good analogies when the analogies are isomorphic to the test problems, but was not beneficial when the analogies are more inclusive than the test problems. In Experiment 2 students were able to effectively combine information from two analogous problems but did significantly worse when combining information form one example and a set of procedures. The last three experiments required that students categorize motion problems according to whether the two distances in a problem should be equated, added, or subtracted. Categorization significantly improved as the number of training examples representing a category increased from one to four (Experiment 3). Categorization was also significantly better when students received both specific and general descriptions of the examples than when they received only a single description (Experiment 4). However, as shown in Experiment 5, students were unable to form their own general descriptions by comparing similar examples.

INTRODUCTION

The objective of this project is to develop a schema-based theory of problem solving to account for how students solve algebra word problems. We are particularly interested in how instructors can effectively organize knowledge to help students become better problem solvers. The research summarized in this report is concerned with the following questions:

(1) What variables influence students' ability to select good analogies?

(2) What kind of information can supplement analogies to make them more effective?

(3) What kind of training is most effective in helping students categorize problems according to common principles?

The first two questions were raised in my 1988 report and are extended here. The third question on categorization represents a new line of research.

The first question is concerned with whether students can evaluate the potential usefulness of an analogous solution. We have been studying how the inclusiveness of an example influences its usefulness and have found that examples which are less inclusive than the test problem are less effective than examples which are more inclusive or isomorphic to the test problem (Reed, Ackinclose, & Voss, in press). Experiment 1 in this report compares students who are mathematics majors with students who have had very few mathematics courses, to determine whether the more experienced students are better at selecting good analogies.

The second question focuses on information that can be used along with an example to make it easier to apply when the example doesn't exactly match the test problem. Our previous research investigated whether students could effectively apply a set of rules to supplement the example and found that students had difficulty with the rules. Experiment 2 extends this line of research by comparing the instructional effectiveness of two examples with the instructional effectiveness of a single example and set of rules. The single example was either simple or complex. The complex example is a new condition that extends the generality of our model of how students combine examples, procedures, and general knowledge to solve test problems that vary in their similarity to the example (Reed & Actor, 1989).

The third question on categorization was studied by asking students to classify motion problems according to whether the two distances in a problem should be equated, added, or subtracted. Experiment 3 investigated how the number of examples influenced students' performance on the posttest. Experiment 4 investigated whether general descriptions of the examples would facilitate categorization of the test problems. Experiment 5 investigated whether students could generate their own general descriptions through the comparison of specific problems, as found by Gick and Holyoak (1983) for Duncker's radiation problem.

SELECTING ANALOGOUS PROBLEMS

In a recent paper on analogical problem solving, Holyoak and Koh (1987) identified four basic steps in transferring knowledge from a source domain to a target domain: (1) constructing mental representations of the source and the target; (2) selecting the source as a potentially relevant analogue to the target, (3) mapping the components of the source and target; and (4) extending the mapping to generate a solution to the target. They stated that the second step, selecting a source analog, is perhaps the least understood of the four steps.

The objective of our study was to identify variables that influence the selection of analogous solutions and to determine whether students would select effective solutions. In the first experiment students had to choose between two problems that belonged to the same category as the test problem. One problem was less inclusive than the test problem and the other problem was more inclusive than the test problem. In the second experiment students had to choose between a problem that was less inclusive than the test problem and a problem that was isomorphic to the test problem.

The same pattern of results occurred in both experiments: students selected problems on the basis of perceived similarity. They did not show a significant preference for the more inclusive problems in the first experiment or the isomorphic problems in the second experiment although both sets of solutions were significantly more effective than solutions to the less inclusive problems. The results therefore reveal a discrepancy between the variable that determines the selection of solutions (similarity) and the variable that determines the usefulness of solutions.

The purpose of the third experiment in this study, summarized in my 1988 AFOSR report, was to determine whether either mathematical experience or the opportunity to study the solutions of analogous problems would increase students' ability to select good analogies. The subjects in the first two experiments were tested in college algebra classes and therefore had similar preparation in mathematics. In contrast, the subjects in the third experiment were participants in the psychology subject pool and therefore had a more varied background in college mathematics courses. They were classified according to three levels of experience, depending on whether they (1) had not taken college algebra, (2) had taken college algebra, or (3) had taken calculus. The second factor -- the opportunity to study the solutions -- was varied by allowing subjects to study the solutions to half of the problem sets before they made their selections.

The results showed that neither more experience nor seeing the solutions increased the selection of more-inclusive solutions. Experience also did not influence the selection of isomorphic solutions, although seeing the solutions did significantly increase the selection of isomorphs.

The failure to find an effect of mathematical experience on selecting solutions may have been caused by an insufficient range in experience. In the present experiment, we included a group of undergraduates who were majoring in mathematics and planned to teach mathematics at a junior high or high school. They were all enrolled in an upper-division mathematics course, Basic Mathematic Concepts, and had previously taken an average of 6 mathematics courses.

Experiment 1: Effect of Experience

Method

Subjects. The subjects were 76 undergraduates at San Diego State University, including the 28 students who were majoring in mathematics. The remaining 48 students were currently taking either an introductory or a cognitive psychology course and received course credit for participating. This group included 29 students who had not taken college algebra (or more advanced courses) and 19 students had taken a college algebra course. Three of the 19 students had also taken a calculus course. All subjects were tested in groups.

Procedure. The instructions indicated that the purpose of the experiment was to determine how people selected related problems to help them solve problems. Students received three sets of questions on similar problems that included a choice between an analogous problem that was less inclusive and an analogous problem that was more inclusive than the test problem. In addition, they received three sets of questions on isomorphic problems that included a choice between an analogous problem that was less inclusive and an analogous problem that was isomorphic to the test problem. Approximately half of the subjects at each level of experience received the similar problems first and the remainder received the isomorphic problems first. This allowed us to evaluate whether presentation order would influence selections.

Results

We analyzed subjects' selection in a 3 (experience) x 2 (presentation order) analysis of variance to determine whether either of these variables would influence the selection of the more inclusive solution. The selections for the 3 similar problems and the 3 isomorphic problems were separately analyzed.

The analysis for the similar problems revealed that neither experience, $F(2, 70) < 1$, nor presentation order, $F(1, 70) < 1$, influenced subjects' preferences. The interaction was also nonsignificant, $F(2, 70) = 2.13$, $MSe = 0.69$ for all tests. The more inclusive solution was selected on 38% of the occasions for students who had not taken college algebra, 39% for students who had taken college algebra, and 42% for students who were mathematics majors.

In contrast, mathematical experience did have a significant effect on the selection of isomorphic problems, $F(2, 70) = 4.37$, $MSe = .59$, $p < .02$. The isomorphic solution was selected on 37% of the occasions for students who had not taken college algebra, 28% of the occasions for students who had taken college algebra, and 50% of the occasions for students majoring in mathematics. Neither presentation order, $F(1, 70) = 2.15$, nor its interaction with experience, $F(2, 70) < 1$, was significant.

Discussion

The finding that students majoring in mathematics did better in selecting isomorphic problems is consistent with the research reported last year that showing students the solutions of the analogous problems improved the selection of isomorphs (Reed, 1988). It is noteworthy that although both of these variables influenced the selection of isomorphs, neither of these variables influenced the selection of similar problems. It is apparently difficult to identify a good analogy when neither of the two analogous problems has a solution that is identical to the test problem. The development of training procedures to help students select good analogies may therefore be beneficial for improving performance on test problems.

EXAMPLES AND PROCEDURES

Although the more-inclusive analogies were more helpful than the less-inclusive analogies, even the more-inclusive solutions helped students solve only 1/3 of the test problems. The difficulty that students have in applying similar solutions caused us to investigate whether anlogous solutions could be supplemented with rules or procedures that would suggest how to modify a similar solution to make it fit the test problem. We therefore gave students either a simple example, a set of procedures, or a simple example and a set of procedures (see Reed, 1988, Experiment 3). The procedures were not very helpful, although we were somewhat successful in designing a model of how students combine examples and procedures in an attempt to construct equations for the test problems in Table 1. In the next experiment we extend this study by including additional instructional conditions in order to test the generality of the model and perhaps discover more effective methods of instruction.

Experiment 2: Simple versus Complex Examples

We designed Experiment 2 to modify and expand the instructional conditions reported in Experiment 3 of Reed (1988). We modified the rules in an attempt to improve their limited effectiveness. In addition, we added 3 new instructional conditions; each including an example

Table 1

Test Problems Used in Experiment 2

1. Bob can paint a house in 12 hours and Jim can paint it in 10 hours. How long will it take them to paint a house if they both work together?

2. Susan can sew a dress in 9 hours and Sherry is three times as fast. How long will it take them to sew a dress if they both work together?

3. An expert can complete a technical task in 5 hours but a novice requires 7 hours to do the same task. When they work together, the novice works 2 hours more than the expert. How long does the expert work?

4. Bill can mow his lawn in 4 hours and his son can mow it in 6 hours. How long will it take both to finish mowing the lawn if they have already mowed 1/3 of it?

5. Jack can build a stereo in 8 hours and Bob is four times as fast. When working together to build a stereo, Bob works 1 hour more than Jack. How long does Jack work?

6. Tom can clean a house in 4 hours and Stan is twice as fast. They clean 1/4 of the house in the morning. How long will it take them to finish cleaning if they continue to work together.

7. A carpenter can build a fence in 7 hours and his assistant can build a fence in 10 hours. On the previous day they built 1/2 of the fence. How long will it take the carpenter to finish the fence if he and his assistant work together, but the assistant works for 3 hours more than the carpenter?

8. John can sort a stack of mail in 6 hours and Paul is twice as fast. They both sort 1/5 of the stack before their break. How long will it take John to sort the remainder if he and Paul work together, but Paul works 1 hour longer?

that was equivalent to the most complex test problem (Problem 8 in Table 1). The 3 new conditions were (1) the complex example, (2) the complex example and procedures, and (3) the complex and simple examples.

We included a complex example as one of the instructional conditions to test the generality of the model. Transformations from the simple example in our previous research created more complex problems. Thus the steep generalization gradients could be caused by both increased complexity and increased dissimilarity from the example. When a complex example is used, the transformations produce simpler problems. This should produce a generalization gradient that is less steep than the gradients obtained for the simple example and allow us to apply the model to a different pattern of results.

The instructional condition that included both a simple and a complex example should be interesting for practical and theoretical reasons. The practical reason is that the ineffectiveness of the rules in our previous research requires the search for other approaches. Examples that are equivalent to Problems 1 and 8 in Table 1 span the set of 8 test problems because the information needed to solve each test problem is contained in the two examples. According to the proposed model, students should apply the pattern matching operation to the quantities in the two examples. All 5 quantities can be obtained from the simple example for Problem 1. Problems 2 - 4 can be solved by matching the simple example on 4 quantities and the complex example on 1 quantity. Problems 5 - 7 can be solved by matching the complex example on 4 quantities and the simple example on 1 quantity. And finally, Problem 8 can be solved by matching the complex example on all 5 quantities.

Another issue was forced upon us during the analysis of the results. We discovered that we had unintentionally modified Problem 8 to read that John works 1 hour longer, rather than Paul works 1 hour longer. In the other test problems (as in the complex example), students were asked to find the time of the person who worked less hours when the hours differed. This issue is somewhat analogous to the role of object correspondences studied by Ross (1987) who found that reversing object correspondences caused a significant decrement in substituting the correct values into a formula. We will discuss the impact of this change when presenting the results.

Method

Subjects. The 174 subjects were enrolled in introductory psychology courses at San Diego State University and received course credit for their participation. They were tested in small groups and assigned randomly to the 6 instructional conditions except for the constraint that there would be an equal number of subjects (29) in each condition. The instructional conditions consisted of the 3 conditions used previously (simple example, procedures, simple example and procedures) and 3 new conditions (complex example, complex example and procedures, simple example and complex example).

Procedure. Students were initially given 5 minutes to attempt to construct a correct equation for the simple and complex example. They then studied for 5 minutes the instructional material which consisted of either a single example, procedures, an example and procedures, or two examples. And finally, they were given 20 minutes to construct equations for the 8 test problems.

Results.

Pretest. There was only one correct equation -- for the simple example -- on the pretest.

Effect of Instruction. The data were analyzed in a 2-factor ANOVA in which intructional method was a between-subjects' factor and transformations was a within-subjects' factor. For subjects in the simple example, procedures, simple example & procedures, and simple example & complex example groups, transformations were measured from the simple example. For subjects in the complex example and complex example & procedures groups, transformations were measured from the complex example. Problem 8 was 0 transformations, Problems 5 to 7 were 1 transformation, Problems 2 to 4 were 2 transformations, and Problem 1 was 3 transformations from the complex example.

Significant effects were found for instructional method, $F(5, 168) = 10.96$, $MSe = .386$, $p < .001$, transformations, $F(3, 504) = 44.91$, $MSe = .072$, $p < .001$, and their interaction, $F(15, 504) = 7.65$, $MSe = .072$, $p < .001$. The percentage of correct equations for each of the instructional groups across the 4 transformations was 7% for the procedures, 32% for the complex example, 38% for the simple example, 45% for the complex example and procedures, 47% for the simple example and procedures, and 65% for the simple and complex examples. According to a Neuman-Keuls analysis, the procedures group performed significantly worse than all other groups at the $p < .01$ confidence level. The group which received two examples performed significantly better at the $p < .01$ level than the groups which received either the procedures, simple example, or complex example and significantly better at the $p < .05$ level for the groups which received the procedures and an example. None of the other paired comparisons was significant.

As mentioned previously, we accidently changed one word in Problem 8 which then stated that John, rather than Paul, worked 1 hour longer. In order to compensate for this change, we did a second analysis of the equations for Problem 8 and scored them as correct if they were correct except for interchanging the variables $\underline{h}$ and $\underline{h + 1}$. This was the way the variables were assigned in the complex example and the other test problems. This change had virtually no effect on the 3 instructional groups that did not receive a complex example, but increased the overall score by approximately 5% for those groups which received the complex example. The rescored means were 7% correct for the procedures, 38% correct for the simple example, 38% correct for the complex example, 48% correct for the simple example and procedures, 49% correct for the complex example and procedures, and 70% correct for the two examples. The F-tests and Neuman-Keuls analysis of the rescored data produced the same pattern of results, at the same confidence levels, as reported above.

Predictions of the Model. Because Problem 8 should be 0 transformations from the complex example we used the more lenient criteria of allowing reversed assignments of h and h + 1 when judging the correctness of an equation. Figure 1 shows the generalization gradients for 4 of the instructional groups, where transformations are measured from the simple example. Figure 2 shows the transformations for the other 2 groups, where transformations are measured from the complex example.

The most extensive version of our model contains 6 parameters to account for the 24 data points in Figures 1 and 2. The parameters result from crossing the 3 operations (matching an example, applying a rule, using general knowledge) with 2 levels of complexity. The three operations are described in my previous report (Reed, 1988, Experiment 3). We again used multiple linear regression to estimate the parameters, which accounted for 94% of the variance as measured by the square of the correlation coefficient. The 6 parameters are generating a correct value by either matching the simple example (m = .97), matching the complex example ($\underline{m}$ = .91), applying a rule to a simple transformation (r = .66), applying a rule to a complex transformation ($\underline{r}$ = .56), applying general knowledge to a simple transformation (g = .64), or applying general knowledge to a complex transformation ($\underline{g}$ = .52).

The application of the model to the 6 instructional groups is shown in Table 2. Consistent with Figures 1 and 2, transformations are measured from the simple example (Figure 1) except for the Complex and Complex & Procedures conditions (Figure 2). The model predicts that the probability of generating a correct equation for the Procedures group ranges from .13 for the simple test problem to .08 for the complex test problem, in which 3 of the 5 quantities are complex.

Predictions for the two single-example conditions replace pattern matching by general knowledge for each transformation. The parameters m = .97 and $\underline{g}$ = .52 were used for the Simple condition and $\underline{m}$ = .91 and g = .64 were used for the Complex condition. Notice that transformations from the simple example produce more complex problems, whereas transformations from the complex example produce simpler problems. Correspondingly, the rule application parameters ($\underline{r}$ = .56 and r = .66) replace the general knowledge parameters to predict the generalization gradients for the Simple & Procedures and the Complex & Procedures conditions.

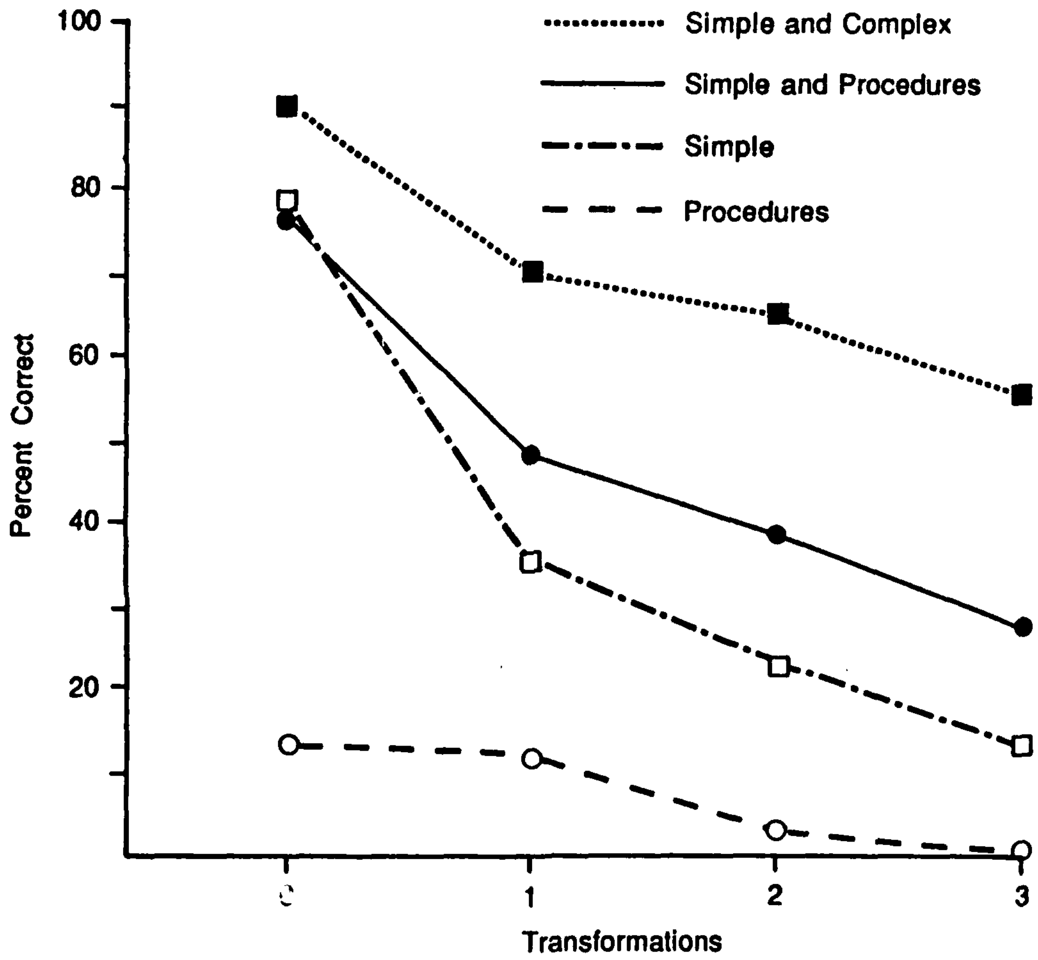

Figure 1. Percent Correct Equations as a Function of Instruction and Transformations from the Simple Example.

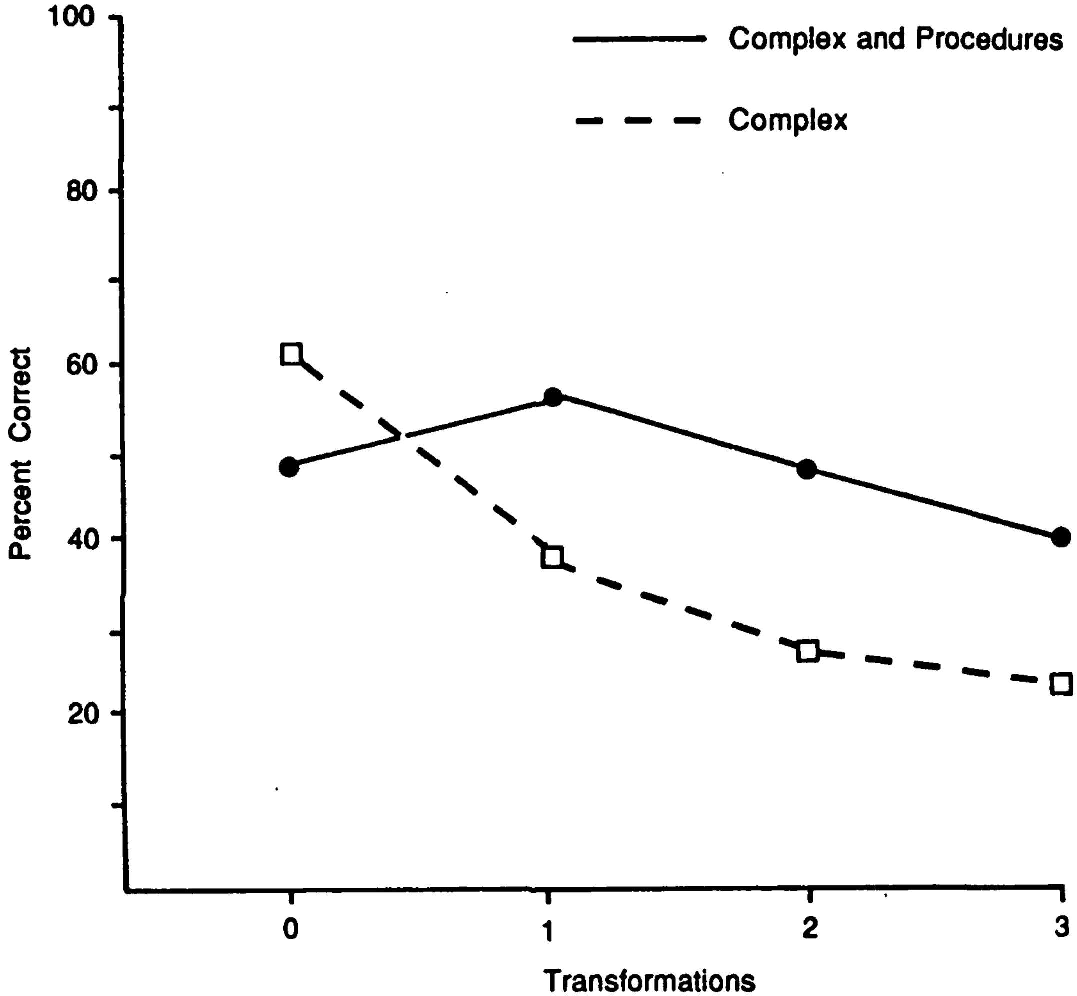

Figure 2. Percent Correct Equations as a Function of Instruction and Transformations from the Complex Example.

Table 2

Observed and Predicted Values for Experiment 2

Transformations	Groups								
	Procedures			**Simple**			**Complex**		
	Observed	Predicted	Model	Observed	Predicted	Model	Observed	Predicted	Model
0	14	13	r^5	79	85	m^5	62	63	$\underline{m}^5$
1	12	11	$r^4\underline{r}$	36	45	$m^4\underline{g}$	37	44	$\underline{m}^4 g$
2	3	9	$r^3\underline{r}^2$	24	24	$m^3\underline{g}^2$	28	31	$\underline{m}^3 g^2$
3	0	8	$r^2\underline{r}^3$	14	13	$m^2\underline{g}^3$	24	21	$\underline{m}^2 g^3$
	Simple and Procedures			**Complex and Procedures**			**Simple and Complex**		
	Observed	Predicted	Model	Observed	Predicted	Model	Observed	Predicted	Model
0	76	85	m^5	48	63	$\underline{m}^5$	90	85	m^5
1	48	49	$m^4\underline{r}$	57	46	$\underline{m}^4 r$	70	79	$m^4\underline{m}$
2	39	29	$m^3\underline{r}^2$	48	33	$\underline{m}^3 r^2$	66	67	$\underline{m}^4 m$
3	28	17	$m^2\underline{r}^3$	41	24	$\underline{m}^2 r^3$	55	63	$\underline{m}^5$

Note: The predictions are based on parameter estimates of m= .97 (matching the simple example), $\underline{m}$= .91 (matching the complex example), r= .66 (applying a rule to a simple transformation), $\underline{r}$= .56 (applying a rule to a complex transformation), g= .64 (applying general knowledge to a simple transformation), and $\underline{g}$= .52 (applying general knowledge to a complex transformation).

Although the model provides fairly accurate predictions for the Simple and the Complex conditions, it predicts gradients that are too steep when the examples are combined with the procedures. The problem is that students perform so poorly when they receive only the procedures that the parameter estimates for correctly applying a procedure are too low to predict the improvement that occurs from having both an example and the set of procedures. In this case, the rule application and general knowledge parameters are sufficiently similar that the number of parameters could be reduced from 6 to 4 without having much effect on the accuracy of the predictions.

The final result is the high level of performance that occurred when students received two examples. The generalization gradient in this case is consistent with the differential performance on the two examples. The model assumes that students use pattern matching to the simple example at 0 transformations and pattern matching to the complex example at 3 transformations. At 1 transformation 4 of the quantities can be obtained from the simple example and 1 quantity can be obtained from the complex example. At 2 transformations (which is 1 transformation from the complex example) 4 of the quantities can be obtained from the complex example and 1 quantity can be obtained from the simple example. The high level of performance across transformations suggests that students are very efficient at performing pattern matching operations, even when they must use more than one example.

CATEGORIZATION

Students in Experiment 2 were given a correct equation and had to learn how to enter appropriate values into the equation. The following experiments on categorization investigate variables that influence students' ability to learn correct equations. A hierarchical model proposed by Reed (1987) claims that constructing an equation for a word problem requires the successful completion of three steps, beginning with the identification of the principle of the problem. For example, motion problems typically require that two distances either be equated, added, or subtracted. Learning the principle requires learning which problems belong in each of these three categories.

There is abundant evidence that the ability to identify the principle of a problem is an important component in acquiring expertise. As people become better problem solvers they group problems together on the basis of common priniciples and mathematical structure rather than on the basis of common objects and story context (Chi, Glaser, & Rees, 1982; Hardiman, Dufresne, & Mestre, 1989; Schoenfeld & Hermann, 1982; Silver, 1981). This finding from sorting studies has been confirmed by the collection of verbal protocols. The explanations generated by good students while studying worked-out examples of mechanics problems demonstrated that they could relate the solution to principles in the text (Chi, Bassock, Lewis, Reimann, & Glaser, 1989). Of particular relevance are the findings obtained by Berger and Wilde (1987) who developed a hierarchical task analysis of algebra word problems that is very similar to my analysis (Reed, 1987). Protocols from their subjects revealed that over half of the experienced students showed a "working-down" strategy that started at the top (principle) level, providing a relatively clear indication of the path to the goal. The novices never started at this level.

The purpose of my study was to investigate how differences in the instructional examples influence the accuracy of classifying motion problems according to the arithmetic operation specified by the principle. Each problem contained two travelled distances that had to be equated, added, or subtracted. The examples were selected from the set of 12 problems shown in Table 3.

The four examples in each category can be further divided into two templates based on Mayer's (1981) taxonomy. According to Mayer, problems belong to the same template if they share the same story line and same list of propositions, regardless of the actual values assigned to each variable or which variable is assigned the unknown. The first and third problems in each category belong to one template (which we will refer to as Template 1) and the second and fourth problems belong to a second template (Template 2). The two examples within a template consist of one problem in which time is the unknown and one problem in which rate is the unknown.

Five of the six templates in Table 3 correspond to templates identified by Mayer. The equate problems consist of overtake and round-trip templates; the addition problems consist of closure and speed-change templates; and the subtraction problems include the same-direction template. The second pair of subtraction problems involves a speed change but the problems specify a difference in two distances rather than a total distance as in the addition problems.

A distinction between Template 1 and Template 2 problems is that all the Template 1 problems consist of a comparison between two objects or people and all the Template 2 problems consist of a single object or person travelling at two different speeds. Although this is a common feature across the three categories, there are also differences. For example, in the overtake problems (equate) the two people travel in the same direction, whereas in the closure problems (add) the two people travel in opposite directions.

Experiment 3: Number of Examples

In Experiment 3 we tested the hypothesis that increasing the number of example problems per category from 1 to 4 would increase accuracy in classifying problems into that category. The hypothesis is based on the assumption that because of limited generalization

Table 3

Instructional Examples for Categorization Experiments

<u>Equate Distances</u>

An escaped convict walks away from a jail at an average speed of 4 mph. A tracker on horseback follows 2 hours later at an average speed of 7 mph. How long will it take the tracker to catch up with the convict?

Distance travelled by convict = Distance travelled by tracker

The Jones family drove to a national park at an average speed of 52 mph and returned along the same route at an average speed of 46 mph. How long did it take to reach the park if it took 1.5 hours longer to return?

Distance to the park = Distance from the park

Jack rides his motorcycle for 2 hours before his brother catches him after riding for 1.5 hours. What was Jack's speed if his brother's speed was 8 mph faster?

Distance travelled by Jack = Distance travelled by Jack's brother

Tim drove to his vacation home in 7 hours and returned the same route in 5 hours. How fast did he drive to his vacation home if his return speed was 18 mph faster?

Distance to vacation home = Distance from vacation home

<u>Add Distances</u>

Howard and Allan live 135 miles apart. They decide to drive toward each other, and Howard drives at 54 mph and Allan drives at 48 mph. How long will Allan drive before they meet if he leaves 1 hour after Howard?

Distance travelled by Howard + Distance travelled by Allan = Total Distance

Susan sailed 45 miles to her favorite island. She sailed 2.5 hours longer at 12 mph than she sailed at 9 mph. How long did she sail at 9 mph?

Distance travelled at 12 mph + Distance travelled at 9 mph = Total Distance

Two trains travel toward each other after leaving citites that are 750 miles apart. The first train travels for 7 hours and the second train travels for 4 hours before they meet. What was the speed of the first train if the second train is 8 mph faster?

Distance travelled by first train + Distance travelled by second train = Total Distance

An athlete trains by running for 1.5 hours and biking for 1 hour, covering a total distance of 25 miles. If his running speed is 10 mph slower than his biking speed, how fast does he run?

Distance travelled running + Distance travelled biking = Total Distance

Table 3 (continued)

<u>Subtract Distances</u>

Tom gave his brother Andy a 0.5 hour head start to see who could return home first. Andy beat his brother by 1 mile by jogging at 3 mph, compared to 4 mph for Tom. How long did it take Andy to reach home?

Distance travelled by Andy - Distance travelled by Tom = Difference in Distances

A man can walk at 1 mph after his accident, compared to 3 mph before his accident. He could walk 2 miles further before his accident by walking 0.5 hours longer at the faster speed. How long can he walk after his accident?

Distance before accident - Distance after accident = Difference in Distances

Karen's boat can travel 150 miles further in 8 hours than Jane's boat can travel in 10 hours. How fast is Karen's boat if it is 25 mph faster than Jane's boat?

Distance for Karen's boat - Distance for Jane's boat = Difference in Distances

A swimmer can swim 2 miles further when she swims for 3 hours with the current than when she swims for 4 hours against the current. If she can swim 1 mph faster with the current, how fast can she swim against the current?

Distance with current - Distance against current = Difference in Distances across examples in a category, students will need to see a variety of examples representing a category. A consequence of the hypothesis is that subjects will be at a disadvantage in classifying templates that are not presented during instruction.

Method

Subjects. The subjects were 50 students enrolled in introductory psychology courses at San Diego State University who received course credit for participating. Ten of the students had taken a college algebra class and 1 student had taken a calculus class. They were tested in small groups and proceeded as a group through the instructional and test material. Two subjects were eliminated because they did not answer any of the questions.

Procedure. The number of instructional examples varied within subjects across the three categories. Each subject received one example for one category, two examples for a second category, and four examples for a third category. The single example was always the first category example in Table 3 and the two examples were either the first and second category examples (representing different templates) or the first and third examples (representing the same template). All four examples in the category were used for the third category. There are 6 ways of assigning 1, 2 or 4 examples to the 3 categories and 2 versions of the 2-example condition, resulting in 12 different instructional booklets. Four subjects were randomly assigned to each instructional booklet.

Subjects began by taking a pretest that consisted of 12 problems that were equivalent to the 12 problems in Table 3. The directions indicated that the 12 problems described a situation in which one or two objects travelled two distances and that the task required identifying whether the two distances should be equated, added, or subtracted. Subjects were told that they should select either Equation 1, 2, or 3 if they thought the distances should be equated, Equation 4, 5, or 6 if they thought the distances should be added, and Equation 7, 8, or 9 if they thought the distances should be subtracted. Selecting a particular equation within a category required distinguishing between whether the distance was stated directly in the problem or whether it must be represented by multiplying rate by time. For example, the three equations for equating the two distances were:

1. Distance 1 = Distance 2
2. Rate x Time = Distance
3. Rate 1 x Time 1 = Rate 2 x Time 2

The purpose of distinguishing among equations within a category was to collect pilot data to aid in the design of subsequent research. Because the instructional examples only illustrated whether subjects should equate, add, or subtract distances, we did not expect much improvement on selecting the correct equation within a category (which was confirmed by the results). We therefore report only whether subjects selected the correct category.

After spending 20 minutes on the pretest, subjects spent 7 minutes studying the 7 examples. The examples appeared on a single page beneath either the label *Equate Distances, Add Distances,* or *Subtract Distances,* with the number of examples (1, 2, or 4) counterbalanced across the three categories. The examples were identical to the ones in Table 3 and included the equations shown below each example.

The posttest consisted of the same 12 problems in the pretest, presented in a different random order. Subjects were not allowed to look at the pretest or the instructional material as they worked on the posttest. The experimenter began collecting the material after 15 minutes, but allowed those subjects who had not finished an additional 5 minutes.

Results

The results were analyzed in an Examples (1, 2, or 4) x Template (Template 1 or Template 2) x Test (pretest or posttest) Anova. The main effect of test was significant, $F(1, 47) = 13.86$, $MSe = 0.35$, $p < .001$, as was the Examples x Test, $F(2, 94) = 4.57$, $MSe = 0.52$, $p < .02$, and Template x Test, $F(1,47) = 4.94$, $MSe = 0.22$, $p < .05$, interactions.

The data supported the hypothesis that increasing the number of examples would increase classification accuracy. Identifying the correct arithmetic operation for those categories that were represented by a single example was as difficult on the posttest (39% correct) as it was on the pretest (41% correct). In contrast, correct identifications increased from 37% to 46% for categories that were represented by two examples, and from 35% to 55% for categories that were represented by four examples.

The Template x Test interaction is consistent with the emphasis on Template 1 examples for two of the four instructional conditions. The category that was represented by only a single (Template 1) example showed a posttest - pretest difference of 7% on Template 1 problems and -11% on Template 2 problems. The category that was represented by two Template 1 examples (the 2-example condition for half of the subjects) showed a 15% gain on Template 1 problems and an 8% gain on Template 2 problems. The category that was represented by one Template 1 and one Template 2 example (the other 2-example condition) showed a 4% gain for Template 1 problems and a 5% gain for Template 2 problems. The category that was represented by two Template 1 problems and two Template 2 problems showed a 24% gain for Template 1 problems and a 17% gain for Template 2 problems. These findings offer moderate support for the hypothesis that learning is template-specific, although the template effect was not as large as expected.

Although the results clearly indicate that the four-example condition produced the biggest increase in correct classifications, it is possible that this increase in the 'hit rate' was achieved by significant increases in the false alarm rate. For instance, if there were four examples illustrating problems in which distances should be equated, subjects might incorrectly classify as equate problems those problems in which the correct operation is to add or subtract. The 20% gain in the correct use of an operation for the four-example condition was accompanied by a 6% increase in the false alarm rate on the posttest. The data indicate that part of the 20% gain in correct classifications was therefore caused by a criterion shift, although the increase in false alarm rate for the four-example condition was nonsignificant, $F(1,47) = 1.89$, $MSe = 1.00$, $p > .05$.

Experiment 4: General Descriptions

Although the examples were effective in improving performance, a more effective procedure might involve presenting both specific examples and more abstract descriptions of the examples. For instance, Cheng, Holyoak, Nisbett, and Oliver (1986) found that training on standard logic was more effective when people received training on both abstract principles of logic and specific examples that illustrated the principles.

The purpose of Experiment 4 was to investigate whether increasing the generality of the example descriptions would affect subjects' ability to identify the correct arithmetic operations. Subjects received either specific examples (as in Experiment 3), a more general description of the examples, or a combination of both specific and general descriptions. The general description summarized the principle for the two templates that represented each of the three categories (see Table 4).

Method

Subjects. The 76 subjects in this experiment were enrolled in either introductory psychology classes (N=58) or a cognitive psychology class (N=18). Eleven students had taken a college algebra course and an additional six students had taken a calculus course.They were tested in small groups and received course credit for their participation.

Procedure. The generality of the examples varied across three instructional conditions. Subjects in the *specific* instructional condition read four specific examples for each mathematical category. The examples were the same ones used in the four-example condition of Experiment 3. Subjects in the *general* instructional condition read two general examples of each mathematical category. The examples provided a more general description of the two templates in each category, as shown in Table 4. Subjects in the *combined* condition received both a general example and a specific example of each template.

Subjects were tested in the same manner as they were in the previous experiment. The experiment began with the pretest, which contained the same 12 test problems used in Experiment 4. Again, subjects were asked to identify whether the distances in each problem should be equated, added or subtracted. This time subjects were only given one choice per mathematical category, instead of the nine categories (three of each) used in Experiment 4.

Table 4

General Descriptions of Principles

Equate Distances

Two distances should be equated when one distance is equal to the other distance. For instance:

1. If two people travel the same route and one overtakes the other, then both travel the same distance. The correct equation shows that the distance travelled by one person equals the distance travelled by the other person.

2. If a person travels the same route on a round trip, then the distance travelled to the destination is the same as the distance travelled from the destination. The correct equation shows that the distance travelled to the destination equals the distance travelled from the destination.

Add Distances

Two travelled distances should be added when one distance plus the other distance combine to form the total distance. For instance:

1. If two people travel toward each other and meet, then the sum of the two travelled distances equals the total distance separating the two people. The correct equation shows that the distance travelled by one person plus the distance travelled by the other person equals the total distance they have to travel.

2. If a person travels at one speed and then changes speed, then the sum of the two travelled distances equals the total distance. The correct equation shows that the distance travelled at one speed plus the distance travelled at the other speed equals the total distance.

Subtract Distances

One travelled distance should be subtracted from the other when it is necessary to use the difference between the two distances. For instance:

1. If one person travels further than another, then the longer distance minus the shorter distance equals the difference in the two distances. The correct equation shows that the shorter distance should be subtracted from the longer distance if one distance is compared to another.

2. If a person travels different distances at different speeds, then the shorter distance should be subtracted from the longer distance if one distance is compared to another. The correct equation shows that the longer distance minus the shorter distance equals the difference in the distances.

Subjects spent 15 minutes on the pretest, 7 minutes on the instruction, and 15 minutes on the posttest.

Results

A 3 (Instruction) x 2 (Test) Anova indicated a significant main effect for test, $F(1, 73) = 41.42$, $MSe = 3.41$, $p < .001$, and a significant interaction between instruction and test, $F(2, 73) = 5.36$, $MSe = 3.41$, $p < .01$. The Instruction x Test interaction supports our hypothesis that the generality of the descriptions would influence correct identifications. Subjects who received both the general and specific examples improved from 34% to 61% in identifying the correct operation. Subjects in the general condition improved from 33% to 43% correct, while subjects in the specific condition improved from 41% to 52% correct. A Newman-Keuls analysis revealed that subjects in the combined condition showed significantly greater improvement than subjects in the general and specific conditions, who did not differ significantly from each other.

These findings are consistent with the results of Cheng et al that presenting both general descriptions and specific examples are more beneficial than presenting either alone. However, the general rules in the Cheng study were very abstract logical rules of the form *If p, then q*. In contrast, the general descriptions in Experiment 4 were domain-specific, identifying round-trip problems, overtake problems, convergence problems, etc, but without the detail of specific examples. The descriptions were therefore at an intermediate level of abstraction and the results support the contention that learning intermediate concepts is useful in constructing a hierarchical knowledge base (Fu & Buchanan, 1985, White, 1989).

Experiment 5: Comparing Descriptions

The results of Experiment 4 revealed that students benefitted from having both a general and a specific description of each template. The purpose of Experiment 5 was to determine whether students could form their own general descriptions through comparing two specific examples of a template. This technique is based on the schema abstraction paradigm developed by Gick and Holyoak (1983). They found that students were more likely to discover the convergence solution to Duncker's radiation problem if they compared two other problems that were solved by using a convergence solution. Gick and Holyoak argued that students formed a more abstract convergence schema through comparing the two specific examples.

We adapted their procedure by asking subjects to describe how two problems were similar to each other. The two problems belonged to the same template (such as two round-trip problems or two overtake problems) for subjects in the same-template group. The two problems belonged to the same-category (such as a round-trip and an overtake problem), but not the same template, for subjects in the different-template group. A third group received instruction that was similar to the instruction received by the combined group in Experiment 4. They received a general description and a specific example and were asked how the general description and specific example are similar to each other.

We anticipated that the combined group would perform the best, based on the results of Experiment 4. However, subjects in the same-template group should approach the performance of the combined group to the extent that they can form general descriptions by comparing two variations of a template. Subjects in the different-template group should perform the worst because their comparisons of different templates should prohibit them from forming general descriptions.

Method

Subjects. The subjects were 98 students enrolled in introductory psychology courses at San Diego State University. They were tested in small groups and received course credit for participating. The vast majority of students had not taken any college mathematics courses, although 8 students had taken a college algebra course and 6 other students had taken a calculus course. Students were randomly assigned to either the same-template group (N = 33), the different-template group (N = 32), or the combined group (N = 33).

Procedure. Subjects received the same pre- and posttest that was given in Experiment 4. They had 15 minutes for each test to decide whether the two distances in the 12 problems shoud be equated, added, or subtracted.

The instruction required that they compare either two problems representing the same template, two problems representing different templates, or a general description and a specific example from the same template. The problems and general descriptions were identical to the ones used in Experiment 4, resulting in two comparisons for each of the three categories.

A separate page of instruction was prepared for each category. The top of the page was labeled either Equate Distances, Add Distances, or Subtract Distances and students were informed that the problems on the page were examples of that category. Students spent 5 minutes on each of the three pages comparing the two pairs of problems that represented the category. The order of presentation was balanced across the three categories.

Results

The results indicated that very little learning occurred. The combined group improved from 34% correct classifications on the pretest to 42% correct on the posttest, the different-template group improved from 39% correct to 43% correct, and the same-template group did not show any improvement (41% vs. 40% correct). Neither the differences between groups, $F(2, 95) < 1$, the differences between tests, $F(1, 95) = 2.74$, $MSe = 3.45$, nor the Group x Test interaction, $F(2, 95) = 1.36$, $MSe = 3.45$, was significant.

The finding that there was not a significant improvement on the posttest is surprising when compared to the highly significant improvement that occurred in Experiments 3 and 4. One general change in procedure that occurred in Experiment 5 was that subjects were not able to simultanously compare problems from contrasting categories. In Experiments 3 and 4 subjects were able to simultanously compare problems from contrasting categories.

A similar finding was reported by Marshall, Barthuli, Brewer, and Rose (1989) who designed a training system to teach students to classify arithmetic story problems into five categories. Training was more successful when all five categories were discussed together than when students received a sequential, detailed description of each category. Her system, Story Problem Solver, therefore begins instruction with an overview and general discussion of all five categories and how they differ from each other.

Another finding that may have limited the performance of the same-template group is that the subjects' similarity descriptions of the two templates rarely mentioned the principles specified in Table 4. Many comparisons were too general, such as *Both involve traveling* or *Both are going from one place to another.* Other comparisons referred to information that was irrelevant to the principle, such as *Both problems are an effort to find distance after worse conditions* or *They both involve people starting at different times.* Still other comparisons mentioned how the distances were combined without specifying a reason, such as *Both require addition* or *They are both subtraction equations.* Only a few of the comparisons specified <u>why</u> the two distances should be added or subtracted or equated.

We are therefore planning an experiment that will ask a more focussed question such as "Why are the two distances added (or subtracted or equated)?" We want to determine whether subjects can provide good reasons and whether their attempts to answer this question result in better performances on the posttest.

References

Berger, D.E., & Wilde, J.M. (1987). A task analysis of algebra word problems. In D.E. Berger, K. Pezdek, & W.P. Banks (Eds.), Applications of Cognitive Psychology: Problem Solving, Education, and Computing, (pp. 123-137). Hillsdale, NJ: Erlbaum.

Cheng, P.W., Holyoak, K.J., Nisbett, R.E., & Oliver,, L.M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. Cognitive Psychology, 18, 293-328.

Chi, M.T.H., Bassock, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13, 145-182.

Chi, M.T.H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R.J. Sternberg (Ed.), Advances in the psychology of human intelligence. (Vol. 1, pp. 7-75). Hillsdale, NJ: Erlbaum.

Fu, L. & Buchanan, B. (1985). Learning intermediate concepts in constructing a hierarchical knowledge base. Proceedings of the Ninth International Joint Conference on Artificial Intelligence, (Vol 1, 559-566).

Gick, M. & Holyoak, K.J. (1983). Schema induction and analogical transfer. Cognitive Psychology, 15, 1-38.

Haridman, P.T., Dufresne, R., & Mestre, J.P. (1989). The relationship between problem categorization and problem solving among experts and novices. Memory & Cognition, 17, 627-638.

Holyoak, K.J. & Koh, K. (1987). Surface and structural similarity in analogical transfer. Memory & Cognition, 15, 332-340.

Marshall, S.P., Barthuli, K.E., Brewer, M.A., & Rose, F.E. (1989). Story problem solver: A schema-based system of instruction. (ONR Tech. Report No. 89-01). San Diego, CA: San Diego State University.

Mayer, R.E. (1981). Frequency norms and structural analyses of algebra story problems into families, categories, and templates. Instructional Science, 10, 135-175.

Reed, S.K. (1987). A structure-mapping model for word problems. Journal of Experimental Psychology: Learning, Memory, & Cognition, 13, 124-139.

Reed, S.K. (1988). Schema-based theories of problem solving. Final Report (Phase 1) to the Air Force Office of Scientific Research, October, 1988.

Reed, S.K., Ackinclose, C.C., & Voss, A.A. (1989). Selecting analogous problems: Similarity versus inclusiveness. Memory & Cognition, in press.

Reed, S.K. & Actor, C. (1989). Use of examples and procedures in problem solving. Manuscript submitted for publication.

Ross, B.H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. Journal of Experimental Psychology: Learning, Memory, & Cognition, 13, 629-639.

Schoenfeld, A.H. & Hermann, D.J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology: Learning, Memory, & Cognition, 5, 484-494.

Silver, E.A. Recall of mathematical problem information: Solving related problems. Journal for Research in Mathematics Education, 1981, 12, 54-64.

White, B.Y. (1989). The role of intermediate abstractions in understanding science and mathematics. Proceedings of the Eleventh Annual Meeting of the Cognitive Science Society. Hillsdate, NJ: Erlbaum.

Publication Information

(1) ' Use of Examples and Procedures in Problem Solving' was submitted to Cognitive Science. Experiment 3 replaced Experiment 1 in the manuscript that I submitted with my 1988 annual report.

(2) 'Constraints on the Abstraction of Solutions' will appear in a special issue of the Journal of Eductional Psychology on mathematical cognition and learning. It is a slight revision of my 1988 version but contains no new research.

(3) 'Selecting Analogous Solutions: Similarity versus Inclusiveness' will appear in Memory & Cognition. Experiment 4 on how mathematical experience influences selections has been added to the 1988 version of this manuscript. In addition, Experiment 3 contains a more thorough analysis of the data.

Presentat

I presented a paper on use of examples and procedures in problem solving on November 10, 1988 in Chicago at the annual meeting of the Psychonomic Society. I also presented colloquia on this topic at the Georgia Institute of Technology, Pomona College, and UCLA.

Personnel

Sanchez, Javier Research Assistant, currently a second-year student in the M.A. program at San Diego State University.